GEOSTATISTICS IN WATER RESOURCES

Shahrokh Rouhani

AUTHORS: School of Civil Engineering, Georgia Institute of Technology, Atlanta, GA 30332.
REFERENCE: *Proceedings of the 1989 Georgia Water Resources Conference*, held May 16 and 17, 1989, at The University of Georgia. Kathryn J. Hatcher, Editor, Institute of Natural Resources, The University of Georgia, Athens, Georgia, 1989.

INTRODUCTION

Geostatistics is a collection of statistical techniques for the analysis of spatial data (Journel and Huijbregts, 1978). In recent years, these tools have developed from research topics into basic techniques in the design and analysis process for a wide range of disciplines, such as mining, geology, and hydrology. The aim of this paper is to present some of the applications of geostatistical tools in water resources.

A typical spatial data set, such as groundwater levels, monthly precipitations, or transmissivities, is composed of scattered readings in space, denoted by z(x), where x represents the measurement location. Having such information, geostatistics provides many techniques to solve a variety of water resources problems, such as:

(i) Estimation of z at an unmeasured location: interpolation and mapping of z;
(ii) Estimation of one variable based on measurements of other variables: co-estimation of piezometric head and transmissivity;
(iii) Estimation of the gradient of z at an arbitrary site: estimation of groundwater flow velocity based on observed heads;
(iv) Estimation of the integral of z over a defined block: estimation of contamination volume based on point measurements; and
(v) Design of sampling and monitoring networks, such as groundwater quality monitoring.

Many of the water related variables are spatial functions presenting complex variations that cannot be effectively described by simple deterministic functions, such as polynomials. Such phenomena are subject of geostatistics that are named as *regionalized variables*. Annual point precipitation is an example of a regionalized variable. Transmissivity also displays spatial variations due to complex processes governing the transport, deposition and compression of materials in sedimentary deposits. Another example of a regionalized variable is the concentration of a chemical compound in groundwater that varies in both space and time. The variations of these processes can be so complicated that estimating their value is difficult, even if measurements from nearby locations are available.

Geostatistics recognizes these difficulties and provides the statistical tools for: (1) calculating the most accurate (according to well-defined criteria) predictions, based on measurements and other relevant information, (2) quantifying the accuracy of these predictions, and (3) selecting the parameters to be measured, and where and when to measure them, if there is an opportunity to collect more data.

Considering that our spatial data represents only an incomplete picture of the natural phenomenon of interest, it is logical to use statistical techniques to process such information. Geostatistics has adopted the language and some of the most practical and yet powerful applicable tools of probability theory.

GEOSTATISTICAL ESTIMATION

On the basis of our limited information, we usually cannot describe our regionalized variable with precision. Instead, it is more convenient to view it as one of many possible *realizations* of a *random field* (a spatial random function). It should be noted here that this modelling assumption is a way to express what is known about our regionalized variable, and does not imply that it is naturally random. Furthermore, it is useful to write

$$Z(x) = m(x) + \xi(x) \qquad (1)$$

where Z(x) = the random variable representing z(x); z(x) = the hydrologic property of interest at x; m(x) = mean value of Z(x) also known as drift or *trend*; $\xi(x)$ = fluctuation value at x, also known as *residual*.

The mean represents a trend that is a deterministic and usually smooth function of space that accounts for the part of Z(x) which represents its large-scale variations. The fluctuation is superimposed on the trend to add up to the point value of Z(x), characterizing the usually small-scale erratic variations of the regionalized variable. This modelling approach enables us to describe unexpected or hard-to-explain variations of a regionalized variable when one moves from one measurement point to another.

The fluctuation term $\xi(x)$ is assumed to oscillate about a zero value and exhibits a statistical structure. The existence of this structure is intuitively justified by the fact that one expects values sampled at neighboring locations would be more alike (numerically) than those values collected at distant locations. To determine this structure based on our observed data, it is imperative to assume a form of stationarity. This stationarity needs to be as weak as possible to make our statistical models more general. An example is the so-called *intrinsic hypothesis* which postulates that: (1) the mean is same everywhere, and (2) for all distances the variance of $Z(x)-Z(x+h)$ is defined and is a unique function of h. In this case, the correlation structure is represented by the *semi-variogram*, defined as

$$\gamma = 1/2 \, \mathrm{Var}[Z(x) - Z(x+h)] \tag{2}$$

where, Var[ ] is the variance. Many authors simply call the above function the variogram. In some cases it is possible to represent the statistical structure by a covariance, which is related to the variogram. Having determined the variogram or the covariance function from our data, we can perform a variety of tasks, including interpolation.

Linear geostatistics estimates the expected value of Z at a location x_0, as a weighted sum of the measured data $z(x_1)$, $z(x_2)$, ..., $z(x_n)$, taken to be realizations of $Z(x)$ at x_1, x_2, ..., x_n, such that

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i \, z(x_i) \tag{3}$$

where $Z^*(x_0)$ = estimated value of Z at x_0; and λ_i's = weights chosen so as to satisfy suitable statistical conditions. The extent of the neighborhood around x_0 determines the n points used in the estimation.

There are two sets of conditions imposed on the above estimated value. The first condition requires that the estimator $Z^*(x_0)$ be unbiased,

$$E[Z^*(x_0) - Z(x_0)] = 0 \tag{4}$$

where $Z(x_0)$ = the value of the random function Z at x_0.

The second condition requires that the estimator $Z^*(x_0)$ have minimum variance of estimation error, which is shown to be

$$V[Z^*(x_0) - Z(x_0)] = 2 \sum_{i=1}^{n} \lambda_i \gamma_{i0} - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma_{ij} \tag{5}$$

where γ_{ij} is $\gamma(|x_i - x_j|)$, variogram between $Z(x_i)$ and $Z(x_j)$.

The minimization of Eq. (5) subject to Eq. (4) is achieved easily by the Lagrangian method, leading to the determination of λ_i's. This form of linear interpolation is referred to as *kriging*, in honor to D.G. Krige, a South African mining engineer who pioneered the use of geostatistics in the assessment of ore bodies. For more information about geostatistical techniques and their applications, readers are referred to the two-part paper of the Task Committee on Geostatistical Techniques (1989a,b).

APPLICATIONS

Geostatistics can be used in a variety of water resources problems, such as:

(i) mapping of spatial variables;
(ii) simulation of hydrological fields;
(iii) co-estimation of hydrological fields using physical relationships, such as co-mapping of piezometric head and transmissivity using groundwater flow equations;
(iv) sampling and monitoring designs; and
(v) water resources system management under uncertainty.

One of the earliest applications of geostatistics in water resources was in the area of mapping spatial variables, such as transmissivity maps, piezometric surfaces, and precipitation fields (Delhomme, 1978). Rouhani (1986) discussed the advantages of geostatistical mapping as compared to other methods, such as least squares or distance weighting. Geostatistical mapping also yields the accuracy map that indicates the areas of high and low precision.

Simulation of hydrologic fields is another application of geostatistics. Simulation usually means the generation of spatial data, such that their mean and their covariance are the same as the original data. There are various useful applications for simulated data. For example, by generating different spatial rainfall patterns, one can determine the statistical distribution of runoff.

Co-estimation allows the user to utilize the information in one variable in the estimation of another. In some instances, a variable that is sampled at a lower cost can be used to improve the accuracy of another variable which is costly to measure. If there are known physical relationships between variables, they can be used to further improve our estimation.

The estimation variance is a measure for the accuracy of estimated fields. This measure can help us to design sampling activities based on the maximization of gained information. Rouhani

(1985) proposes the variance reduction analysis to identify the best locations for further sampling. This method is a useful procedure for planning activities (Rouhani and Fiering, 1986).

In some instances, such as groundwater quality monitoring, the estimated magnitude of the variable of interest is as important as its accuracy. So the sampling may be designed not only for improving the precision of the estimated field, but also for targeting those areas which exhibit critical estimated values. Rouhani and Hall (1988) have developed appropriate techniques for groundwater sampling and applied them to groundwater quality monitoring in southwestern Georgia.

Water resources management problems usually include many variables that exhibit uncertainty. Ignoring the stochastic nature of these problems may yield non-optimal solution. Geostatistics provides the frame work to quantify these uncertainties and incorporate them in our decisions.

ACKNOWLEDGMENT

This research was supported by grant INT-8702264 from the National Science Foundation.

LITERATURE CITED

Delhomme, J.P., 1978. Kriging in the Hydrosciences. Advances in Water Resources, 1(5):251-266.

Journel, A.G., and C.J. Huijbregts, 1978. Mining Geostatistics, Academic Press, London, 600 p.

Rouhani, S., 1985. Variance Reduction Analysis. Water Resources Research, 21(6): 837-846.

Rouhani, S., 1986. Comparative Study of Ground Water Mapping Techniques. Journal of Ground Water, 24(2):207-216.

Rouhani, S., and M.B. Fiering, 1986. Resilience of Statistical Sampling Scheme. Journal of Hydrology, 89:1-11.

Rouhani S., and T.J. Hall, 1988. Geostatistical Schemes for Groundwater Sampling. Journal of Hydrology, 103:85-102.

Task Committee on Geostatistical Techniques in Geohydrology (S. Rouhani, Chairman and Principal Author), 1989a. Review of Geostatistics in Geohydrology, 1. Basic Concepts. ASCE Journal of Hydraulic Engineering, submitted.

Task Committee on Geostatistical Techniques in Geohydrology (S. Rouhani, Chairman and Principal Author), 1989b. Review of Geostatistics in Geohydrology, 2. Applications. ASCE Journal of Hydraulic Engineering, submitted, 1989.